

# Fast variables determine the epidemic threshold in the pairwise model with an improved closure

István Z. Kiss, Joel C. Miller and Péter L. Simon

**Abstract** Pairwise models are used widely to model epidemic spread on networks. These include the modelling of susceptible-infected-removed (SIR) epidemics on regular networks and extensions to SIS dynamics and contact tracing on more exotic networks exhibiting degree heterogeneity, directed and/or weighted links and clustering. However, extra features of the disease dynamics or of the network lead to an increase in system size and analytical tractability becomes problematic. Various “closures” can be used to keep the system tractable. Focusing on SIR epidemics on regular but clustered networks, we show that even for the most complex closure we can determine the epidemic threshold as an asymptotic expansion in terms of the clustering coefficient. We do this by exploiting the presence of a system of fast variables, specified by the correlation structure of the epidemic, whose steady state determines the epidemic threshold. While we do not find the steady state analytically, we create an elegant asymptotic expansion of it. We validate this new threshold by comparing it to the numerical solution of the full system and find excellent agreement over a wide range of values of the clustering coefficient, transmission rate and average degree of the network. The technique carries over to pairwise models with other closures [1] and we note that the epidemic threshold will be model dependent. This emphasises the importance of model choice when dealing with realistic outbreaks.

---

István Z. Kiss

School of Mathematical and Physical Sciences, Department of Mathematics, University of Sussex, Falmer, Brighton BN1 9QH, UK, e-mail: [i.z.kiss@sussex.ac.uk](mailto:i.z.kiss@sussex.ac.uk)

Joel C. Miller

Institute for Disease Modeling, Bellevue, WA, United States of America e-mail: [joel.c.miller.research@gmail.com](mailto:joel.c.miller.research@gmail.com)

Péter L. Simon

Institute of Mathematics, Eötvös Loránd University Budapest, and Numerical Analysis and Large Networks Research Group, Hungarian Academy of Sciences, Hungary, e-mail: [simonp@math.elte.hu](mailto:simonp@math.elte.hu)

## 1 Introduction

One way to deal with the challenges of modelling stochastic epidemics on networks is to use mean-field models. This approach has led to a number of models including heterogeneous or degree-based mean-field [18, 17], pairwise [19, 8], effective-degree [11], edge-based compartmental [14] and message passing [6], to name a few. The main difference between these models is how the variables over which averaging is done are chosen. Perhaps the most compact model is the edge-based compartmental model [15] and this works for heterogeneous networks with Markovian SIR epidemics, although extensions of it for arbitrary infection and recovery processes are also possible [23].

Pairwise models are popular and the first model for regular networks and SIR epidemics [19, 8] was generalised to heterogeneous networks [3], preferentially mixing networks [3], directed [22] and weighted networks [20], adaptive networks [9], and structured networks [4] among others. Its wide use is perhaps due to its relative transparency where variables are defined in a straightforward way. A downside of the pairwise models is that in constructing them we find that the change in the expected number of individual nodes of a given state depends on to the expected number of edges (or pairs) between nodes of various states. The change in the expected number of edges depends on larger-scale structure. To keep the system tractable, we generally make a “closure assumption” that we can express the frequency of the relevant larger-scale structures in terms of the pairs and individuals, that is lower order moments or structure.

A basic understanding of the network and epidemic dynamics coupled with careful bookkeeping and an appropriate closure assumption produces a pairwise model. Pairwise models have been successfully used to analytically derive the epidemic threshold and final epidemic size. However, these results are mostly limited to networks without clustering. The propensity of contacts to cluster, i.e. that two friends of an individual/node are also friends of each other, is known to lead to many complications, and modelling epidemics on clustered networks using analytically tractable mean-field models is still limited to networks with very specific structural features [4, 16, 12, 13, 7, 25, 21]. However, using approaches borrowed from percolation theory [13] and focusing more on the stochastic process itself [24], some results have been obtained.

For pairwise models, clustering first manifests itself by requiring a different and more complex closure, which makes the analysis of the resulting system, even for regular networks and SIR dynamics, challenging. Furthermore, it turns out that such closures may in fact fail to conserve pair-level relations and may not accurately reflect the early growth of quantities such as closed loops of three nodes with all nodes being infected [5]. Such considerations have led to an improved closure being developed in an effort to keep as many true features of the exact epidemic process as possible [5]. In this paper we will focus on the classic pairwise model for regular networks with clustering but using the improved closure of [5], given below in equation (8). We will show that by working with fast variables corresponding to the correlations that develop during the spread of the epidemic, we can analytically de-

termine the epidemic threshold as an asymptotic expansion in terms of the clustering coefficient.

The use of fast variables is not completely new. They were used in [8] and [2] but not with the improved closure. Even with the simpler closures, the epidemic threshold has only been obtained numerically and it was framed in terms of a growth-rate-based threshold (which is equivalent to the basic reproduction number at the critical point of the epidemic spread). In [2] a hybrid pairwise model incorporating random and clustered contacts is considered, with the analysis focused on the growth-rate-based threshold. The authors of [2] managed to derive a number of results, some analytic (the critical clustering coefficient for which an epidemic can take off) and some semi-analytic, and they have shown, in agreement with most studies, that clustering inhibits the spread of the epidemic when compared to an equivalent network without clustering but with equivalent parameter values governing the epidemic process. However, no analytic expression for the threshold was provided. More recently, in [10], the epidemic threshold in a pairwise model for clustered networks with closures based on the number of links in a motif, rather than nodes, was calculated.

Building on these results and the recent paper by Barnard et al [1] (where the idea of fast variables was used to derive an analytic epidemic threshold for pairwise models with two different closures corresponding to clustered networks) we set out to take the final step of using fast variables and perturbation theory to determine an asymptotic expansion of the epidemic threshold when the pairwise model is closed with equation (8). The paper is structured as follows. In Section 2 we outline the model. The main results, both analytical and numerical, are presented in Section 3. We conclude with a discussion of the results and possible extensions in Section 4.

## 2 Model formulation

### 2.1 The network and standard SIR dynamics

We begin by considering a population of  $N$  individuals and describe their contact structure by an undirected network with adjacency matrix  $G = (g_{ij})_{i,j=1,2,\dots,N}$  where  $g_{ij} = 1$  if nodes  $i$  and  $j$  are connected and zero otherwise. Because the network is undirected,  $g_{ij} = g_{ji}$  for all  $i, j = 1, 2, \dots, N$ , and because we exclude self-loops,  $g_{ii} = 0$  for all  $i$ . The network is static and regular, such that each individual has exactly  $n$  edges or links. The sum over all elements of  $G$  is defined as  $\|G\| = \sum_{i,j} g_{ij}$ . Hence, the number of doubly counted links in the network is  $\|G\| = nN$ . More importantly, using simple matrix operations on  $G$ , we can calculate the clustering coefficient of the network

$$\phi = \frac{\text{trace}(G^3)}{\|G^2\| - \text{trace}(G^2)}, \quad (1)$$

where  $\text{trace}(G^3)$  yields six times the number of closed triples or loops of length three (uniquely counted) and  $||G^2|| - \text{trace}(G^2)$ , twice the number of triples (open and closed, also uniquely counted).

Let us consider the standard SIR epidemic dynamics on a network. The dynamics are driven by two processes: (a) infection and (b) recovery from infection. Infection can spread from an infected/infectious node to any of its susceptible neighbours. We model this as a Poisson point process with per-link infection rate  $\tau$ . Infectious nodes recover at constant rate  $\gamma$ , independently of the network, and gain permanent immunity.

## 2.2 The unclosed pairwise model

Let  $A_i$  be 1 if the individual at node  $i$  is of type  $A \in \{S, I, R\}$  and zero otherwise. Then single nodes (singles) of type  $A$  can be counted as  $[A] = \sum_i A_i$ , pairs of nodes (pairs) of type  $A - B$  can be counted as  $[AB] = \sum_{i,j} A_i B_j g_{ij}$  and triples of nodes (triples) of type  $A - B - C$  can be counted as  $[ABC] = \sum_{i,j,k} A_i B_j C_k g_{ij} g_{jk}$ . This method of counting means that pairs are counted once in each direction, so  $[AB] = [BA]$ , and  $[AA]$  is even. Using this notation to track singles, pairs, and triples leads to the following system of pairwise equations describing the SIR epidemic on a regular network:

$$[\dot{S}] = -\tau[SI]; \quad [\dot{I}] = \tau[SI] - \gamma[I]; \quad [\dot{R}] = \gamma[I], \quad (2)$$

$$[\dot{SI}] = \tau([SSI] - [ISI] - [SI]) - \gamma[SI], \quad (3)$$

$$[\dot{SR}] = \gamma[SI] - \tau[RSI], \quad (4)$$

$$[\dot{II}] = 2\tau([ISI] + [SI]) - 2\gamma[II], \quad (5)$$

$$[\dot{IR}] = \gamma([II] - [IR]) + \tau[RSI]. \quad (6)$$

We note that equations (3)–(6) contain triples which are not defined within the entire system of equations (2)–(6). Furthermore, we have chose these variables in order to be able to consistently define our fast variables later. To determine solutions of the system, we must find a way to account for these triples in terms of pairs and singles through a closure assumption. It is worth noting that this system is exact before a closure is implemented [9].

## 2.3 The improved closure and the closed pairwise system

The key for deriving the improved closure [5] is to split the non-clustered and clustered part of the network and to determine the propensity of a susceptible node's neighbour to be in state  $A$  (where  $A \in \{S, I, R\}$ ), given that the susceptible node is already connected to an infected one. This can be defined as

$$p_{A|S-I} = \begin{cases} p_{A|S-I}^{uc} & \text{with probability } (1 - \phi), \\ \frac{p_{A|S-I}^c}{\sum_a p_{a|S-I}^c} & \text{with probability } \phi, \end{cases} \quad (7)$$

where  $p_{A|S-I}^{uc} = \frac{[AS]}{n[S]}$ ,  $p_{A|S-I}^c = p_{A|S-I}^{uc} C_{AI}$  and  $C_{AI} = \frac{N[A]I}{n[A][I]}$ . In the absence of clustering we assume that the probability the neighbour is of state  $A$  is simply given by frequency of  $[AS]$  type links relative to all links emanating from susceptible nodes,  $n[S]$ . If clustering is present then the probability of finding a susceptible neighbour decreases as the transitive link connects this particular neighbour to the existing infected neighbour. This means that the node is exposed to infection and its probability of remaining susceptible decreases. This effect is captured by  $C_{AI}$  which expresses how much more probable it is, compared to the random mixing case, to find a neighbour in state  $A$  given that the node is also connected to an infectious node. It is well known that epidemics are negatively correlated in the sense that we are more likely to find  $I-I$  type links rather than  $I-S$ . Unfortunately,  $p_{A|S-I}^c$  alone is not a properly defined probability. Despite this the closure resulting from it has been used although it leads to some anomalies such as non-conservation of pair-level relations. However, the normalised form of it, as in equation (7), leads to the improved closure [5]. Taking into account the new way of defining  $p_{A|S-I}$ , this yields

$$\begin{aligned} [ASI]_c &= (1 - \phi)[ASI] + \phi[ASI] = (1 - \phi)(n - 1)[SI]p_{A|S-I}^{uc} + \phi(n - 1)[SI]\frac{p_{A|S-I}^c}{\sum_a p_{a|S-I}^c} \\ &= (1 - \phi)(n - 1)[SI]\frac{[AS]}{n[S]} + \phi(n - 1)[SI]\frac{\frac{[AS]}{n[S]}C_{AI}}{\sum_a p_{a|S-I}^c} \\ &= (1 - \phi)\frac{(n - 1)}{n}\frac{[AS][SI]}{[S]} + \phi(n - 1)[SI]\frac{\frac{[AS]}{n[S]}\frac{N[A]I}{n[A][I]}}{\sum_a \frac{[aS]}{n[S]}\frac{N[a]I}{n[a][I]}} \\ &= (1 - \phi)\frac{(n - 1)}{n}\frac{[AS][SI]}{[S]} + \phi(n - 1)\frac{[AS][SI][IA]}{[A]\sum_a \frac{[aS][aI]}{[a]}} \\ &= (n - 1)\left((1 - \phi)\frac{[AS][SI]}{n[S]} + \phi\frac{[AS][SI][IA]}{[A]\sum_a [aS][aI]/[a]}\right), \end{aligned} \quad (8)$$

where  $a \in \{S, I, R\}$  and  $[ASI]_c$  is used to distinguish this approximation from its exact equivalent.

### 3 Results for the pairwise model with the improved closure

Plugging equation (8) into the exact system (2)–(6) leads to the self-consistent system below

$$[\dot{S}] = -\tau[SI]; \quad [\dot{I}] = \tau[SI] - \gamma[I]; \quad [\dot{R}] = \gamma[I] \quad (9)$$

$$[\dot{SI}] = -(\tau + \gamma)[SI] + \tau[SSI]_c - \tau[ISI]_c, \quad (10)$$

$$[\dot{SR}] = \gamma[SI] - \tau[RSI]_c, \quad (11)$$

$$[\dot{II}] = 2\tau[SI] - 2\gamma[II] + 2\tau[ISI]_c, \quad (12)$$

$$[\dot{IR}] = \gamma([II] - [IR]) + \tau[RSI]_c, \quad (13)$$

where  $[ASI]_c$  with  $A \in \{S, I, R\}$  is defined in equation (8). The standard linear stability analysis of this system around the disease free steady state,  $([S], [I], [SI], [SS], [II]) = (N, 0, 0, nN, 0)$  leads to some terms such as

$$\alpha = \frac{[SI]}{[I]}, \delta = \frac{[II]}{[I]}, x = \frac{[SR]}{[R]}, y = \frac{[IR]}{[I]}. \quad (14)$$

Interestingly these terms are ill-defined since both denominators and numerators are zero at the equilibrium. However, these variables have a clear biological meaning and are related to the correlation structure of the epidemic.

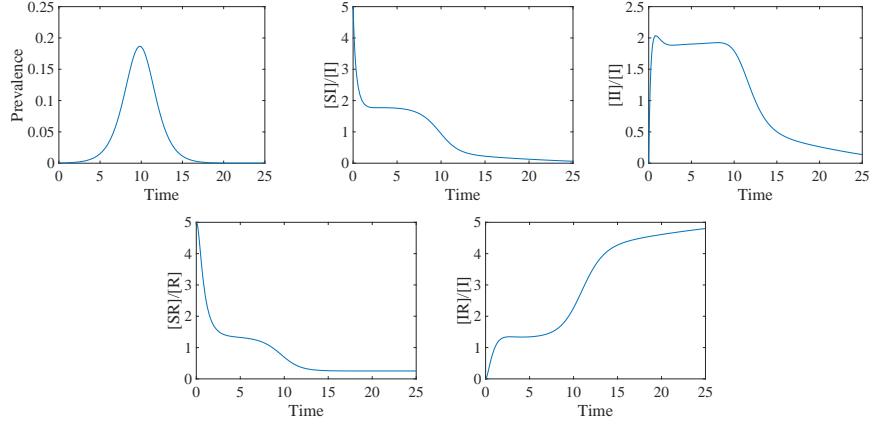
Interestingly however, the epidemic threshold can also be found in a more direct way by looking at equation (9). Namely, this leads to

$$[\dot{I}] = \tau[SI] - \gamma[I] = \gamma[I] \left( \frac{\tau}{\gamma} \frac{[SI]}{[I]} - 1 \right), \quad (15)$$

which clearly shows that the epidemic threshold coincides with  $\overline{\mathcal{R}}_0 = \frac{\tau}{\gamma} \frac{[SI]}{[I]} = 1$ . This is a growth-rate-based threshold of the epidemic and while  $\overline{\mathcal{R}}_0$  is different from the basic reproduction number, they are equivalent when both are exactly one. From here, we can see that finding the threshold amounts to finding  $\alpha = \frac{[SI]}{[I]}$  at time  $t$  close to zero. As we will show next, these new variables of interest are fast variables and settle quickly, even if only temporarily, to a quasi-equilibrium. The time taken to reach this quasi-equilibrium is short compared to the timescale of epidemic growth, and the quasi-equilibrium corresponds to the exponential growth phase of the epidemic.

### 3.1 Fast variables with the improved closure

In Fig. 1 we plot the evolution of the prevalence together with the newly defined variables. The figure shows clearly that these new variables are fast. Namely, early on, when the prevalence is small, these variables settle to a temporary equilibrium. The natural step would be to investigate the behaviour of  $\alpha$  (and all the others) and this can be done by deriving their evolution equations, i.e. compute  $d\alpha/dt = d([SI]/[I])/dt$  and use equations (9)–(13). After some simple but lengthy algebra one obtains



**Fig. 1** Time evolution of the prevalence and four fast variables based on numerical integration of the closed pairwise system (9)–(13). Parameter values are  $N = 10000$ ,  $\gamma = 1$  and  $\phi = 0.5$ .

$$\frac{d\alpha}{dt} = -\tau\alpha(1 + \alpha) + \tau\xi(1 - \phi)n\alpha + \tau(n - 1)\phi \left( \frac{n\alpha^2 - \alpha^2\delta}{n\alpha + \alpha\delta + xy} \right). \quad (16)$$

The assumption of being close to  $t = 0$  is used to neglect a term of the form  $\tau\xi(1 - \phi)\frac{[SI]}{[I]}\frac{[SI]}{[S]}$ , together with a few others with a similar structure. While  $\alpha = [SI]/[I]$  itself is a well-defined and bounded ratio of two small numbers,  $[SI]/[S] \simeq 0$  when  $t$  is close to zero. We use similar arguments when deriving the equations for the other variables. Their differential equations are

$$\frac{d\delta}{dt} = -\gamma\delta + 2\tau\alpha - \tau\alpha\delta + 2\tau(n - 1)\phi \frac{\alpha^2\delta}{n\alpha + \alpha\delta + xy}, \quad (17)$$

$$\frac{dx}{dt} = +\tau\alpha^*(\alpha - x) - \gamma(\alpha - x) - \frac{\tau(n - 1)\phi\alpha xy \left( \frac{\tau}{\gamma}\alpha^* - 1 \right)}{n\alpha + \alpha\delta + xy}, \quad (18)$$

$$\frac{dy}{dt} = +\tau\delta - \tau\alpha y + \frac{\tau(n - 1)\phi\alpha xy}{n\alpha + \alpha\delta + xy}. \quad (19)$$

As one notices the four variables are interlinked and are all needed to resolve the evolution equation of each. A key step in the derivation above is the need to introduce  $\alpha^*$  which corresponds to the steady state of the system defined by equations (16)–(19). This is needed as in the derivation of the evolution equations for  $x$  terms such as  $[SI]/[R]$ ,  $[IR]/[R]$  and  $[I]/[R]$  can only be dealt with by noticing that at time  $t$  close to  $t = 0$  we have that

$$[I] = \left( \frac{\tau}{\gamma}\alpha^* - 1 \right) [R]. \quad (20)$$

This follows from the assumption that  $[I] \approx ce^{rt}$  for some  $c$  and  $r$  and  $d[R]/dt = \gamma I$ , or from the observation that

$$\frac{d[I]}{d[R]} = \frac{\tau[SI] - \gamma[I]}{\gamma[I]} = \left( \frac{\tau}{\gamma} \frac{[SI]}{[I]} - 1 \right) \simeq \left( \frac{\tau}{\gamma} \alpha^* - 1 \right), \quad (21)$$

where we assumed that  $\alpha$  stabilises quickly at small time. Integrating this leads to  $[I] = \left( \frac{\tau}{\gamma} \alpha^* - 1 \right) [R] + C$ , where  $C = 0$  if the initial conditions at  $t = 0$  are plugged in. This in turn allows us to write  $[SI]/[R] = ([SI]/[I])([I]/[R])$  and  $[IR]/[R] = ([IR]/[I])([I]/[R])$  which ensures that we can cast all terms as functions of the four fast variables.

### 3.2 Asymptotic expansion of the epidemic threshold

Finding the steady state of the system defined by equations (16)–(19) may seem like a difficult task but it turns out that an asymptotic solution is within reach. To do this each variable  $v$  is written as  $v = v_0 + \phi v_1 + \dots$ , where  $v \in \{\alpha, \delta, x, y\}$ . Plugging these into equations (16)–(19) leads to the following system at  $\mathcal{O}(1)$ :

$$((n-2) - \alpha_0)(n\alpha_0 + \alpha_0\delta_0 + x_0y_0) = 0, \quad (22)$$

$$(-\gamma\delta_0 + 2\tau\alpha_0 - \tau\alpha_0\delta_0)(n\alpha_0 + \alpha_0\delta_0 + x_0y_0) = 0, \quad (23)$$

$$(\delta_0 - \alpha_0y_0)(n\alpha_0 + \alpha_0\delta_0 + x_0y_0) = 0, \quad (24)$$

$$(\tau\gamma\alpha_0^2 - \tau\gamma\alpha_0x_0 - \gamma^2\alpha_0 + \gamma^2x_0)(n\alpha_0 + \alpha_0\delta_0 + x_0y_0) = 0. \quad (25)$$

One of the solutions of the system above is:

$$(\alpha_0, \delta_0, x_0, y_0) = \left( n-2, \frac{2\tau(n-2)}{\gamma + \tau(n-2)}, n-2, \frac{2\tau}{\gamma + \tau(n-2)} \right). \quad (26)$$

At  $\mathcal{O}(\phi)$  from equation (16) we have

$$-(\alpha_1 + (n-1))(n\alpha_0 + \alpha_0\delta_0 + x_0y_0) + (n-1)\alpha_0(n - \delta_0) = 0. \quad (27)$$

Plugging in the solutions at  $\mathcal{O}(1)$  (see eq. (26)) into the equation above leads to

$$\alpha_1 = -\frac{2\tau(2n-3)(n-1)}{n(\gamma + \tau(n-2)) + 2\tau(n-1)}. \quad (28)$$

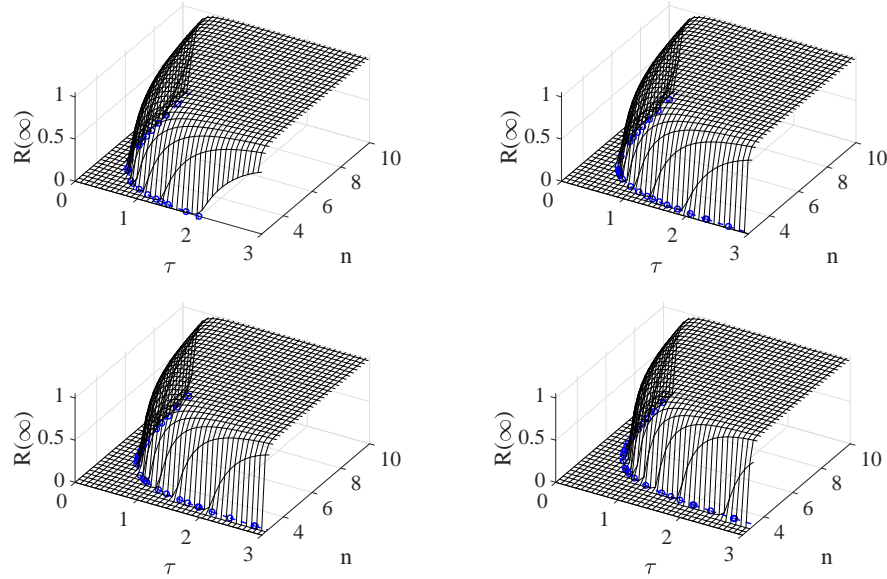
Hence the epidemic threshold, up to the first correction is given by  $\overline{\mathcal{R}}_0 = 1$  where

$$\overline{\mathcal{R}}_0 = \frac{\tau}{\gamma}(\alpha_0 + \phi\alpha_1) = \frac{\tau(n-2)}{\gamma} - \frac{\tau}{\gamma} \frac{2\tau(2n-3)(n-1)}{n(\gamma + \tau(n-2)) + 2\tau(n-1)}\phi. \quad (29)$$



The first observation that can be made is that the first order correction is negative and this implies that clustering reduces the epidemic threshold and makes the epidemic less likely to spread. The second is that when  $\phi = 0$ ,  $\mathcal{R}_0 = 1$  reduces to the well known threshold when a network with no clustering is considered.

### 3.3 Numerical examples



**Fig. 2** Assessing the validity of the epidemic threshold based on the asymptotic expansion (29) (dashed line and markers -  $\circ$ ) by comparing it to the threshold based on the numerical solution of closed pairwise system (9)–(13) (surface). Parameter values are  $N = 10000$ ,  $\gamma = 1$  and the clustering coefficients are top (from left to right)  $\phi = 0, 0.2$ , and bottom (from left to right)  $\phi = 0.4, 0.6$ .

In Fig. 2 we show a systematic test of comparing the epidemic threshold generated via solving the closed pairwise system (9)–(13) numerically to the epidemic threshold based on the asymptotic expansion (29), over a wide range of  $(\tau, n)$  values. Several observations can be made. First, it is clear that higher values of clustering push the location of threshold to higher  $\tau$  and  $n$  values, meaning that the limiting effect of clustering on the epidemic spread can only be overcome if either the value of the transmission rate or average degree increases. Second, the agreement between the numerical and asymptotic threshold is excellent for a large range of clustering values. In fact, a slight discrepancy only really seems to appear at around  $\phi = 0.6$ .

It is worth noting that finding the final epidemic size numerically can be achieved by using a more compact system. However, the extended system is preferred here since the derivation of the system of fast variables relies upon it.

## 4 Discussion

In this paper we set out to obtain an analytic epidemic threshold using the pairwise model an improved closure to account for clustering. This problem has been solved previously in the unclustered case [8]. Here, we went one step further and showed that the quasi-equilibrium can be found as an asymptotic expansion in powers of the clustering coefficient. This paper builds on work in [1] and shows that exploiting the presence of fast variables and combining it with perturbation theory leads to a fruitful methodology which allowed us to compute the epidemic threshold analytically from pairwise models with three different closures. Strictly speaking there is no reason why this approach would not apply to other systems with properties similar to those found in the pairwise model. Reflecting on the results in [1] and in the present paper it is obvious that the epidemic threshold is model dependent and care has to be taken if such a model is used to model a real outbreak.

The ODE systems for the fast variables are worth investigating in more detail. We expect that these systems will exhibit a number of steady states. In fact preliminary numerical simulations suggest that the system corresponding to the fast variables (16)–(19) has at least one steady state which is identical to the quasi-steady states shown in Fig. 1. Furthermore, it would be interesting to consider if the idea of fast variables extends to other mean-field models used in epidemiology. In particular it would be worthwhile to investigate if the correlation structure maps onto multi-variable models for heterogenous networks and if this consideration may lead to new insight from existing models. Equally, it remains a challenge to derive compact mean-field models for clustered networks. However, if such models will materialise we expect that our method may be a good candidate when it comes to the analysis of such models.

Finally, the natural next step would be to test our findings against explicit stochastic network simulations. This was beyond the scope of the present work, whose focus was on exploiting the presence of fast variables and the use of perturbation analysis to determine the epidemic threshold analytically.

## Acknowledgments

István Z. Kiss acknowledges support from the Leverhulme Trust Research Project Grant (RPG-2017-370). Péter L. Simon acknowledges support from Hungarian Scientific Research Fund, OTKA, (grant no. 115926). Joel C. Miller acknowledges support from Global Good.

## References

1. Rosanna C Barnard, Luc Berthouze, Péter L Simon, and István Z Kiss. Epidemic threshold in pairwise models for clustered networks: closures and fast correlations. *arXiv preprint arXiv:1806.06135*, 2018.
2. Ken TD Eames. Modelling disease spread through random and regular contacts in clustered populations. *Theoretical population biology*, 73(1):104–111, 2008.
3. Ken TD Eames and Matt J Keeling. Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. *Proceedings of the National Academy of Sciences*, 99(20):13330–13335, 2002.
4. Thomas House, Geoffrey Davies, Leon Danon, and Matt J Keeling. A motif-based approach to network epidemics. *Bulletin of Mathematical Biology*, 71(7):1693–1706, 2009.
5. Thomas House and Matt J Keeling. The impact of contact tracing in clustered populations. *PLoS computational biology*, 6(3):e1000721, 2010.
6. Brian Karrer and Mark EJ Newman. Message passing approach for general epidemic models. *Physical Review E*, 82(1):016101, 2010.
7. Brian Karrer and Mark EJ Newman. Random graphs containing arbitrary distributions of subgraphs. *Physical Review E*, 82(6):066118, 2010.
8. Matthew J Keeling. The effects of local spatial structure on epidemiological invasions. *Proceedings of the Royal Society of London B: Biological Sciences*, 266(1421):859–867, 1999.
9. István Z Kiss, Joel C Miller, and Péter L Simon. *Mathematics of Epidemics on Networks*. Springer, 2017.
10. Jinxian Li, Weiqiang Li, and Zhen Jin. The epidemic model based on the approximation for third-order motifs on networks. *Mathematical biosciences*, 2018.
11. Jennifer Lindquist, Junling Ma, P Van den Driessche, and Frederick H Willeboordse. Effective degree network disease models. *Journal of mathematical biology*, 62(2):143–164, 2011.
12. Joel C Miller. Percolation and epidemics in random clustered networks. *Physical Review E*, 80(2):020901, 2009.
13. Joel C Miller. Spread of infectious disease through clustered populations. *Journal of the Royal Society Interface*, pages rsif–2008, 2009.
14. Joel C Miller, Anja C Slim, and Erik M Volz. Edge-based compartmental modelling for infectious disease spread. *Journal of the Royal Society Interface*, 9(70):890–906, 2012.
15. Joel C Miller and Erik M Volz. Model hierarchies in edge-based compartmental modeling for infectious disease spread. *Journal of mathematical biology*, 67(4):869–899, 2013.
16. Mark EJ Newman. Random graphs with clustering. *Physical review letters*, 103(5):058701, 2009.
17. Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. Epidemic processes in complex networks. *Reviews of modern physics*, 87(3):925, 2015.
18. Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic dynamics and endemic states in complex networks. *Physical Review E*, 63(6):066117, 2001.
19. DA Rand. Correlation equations and pair approximations for spatial ecologies. *Advanced ecological theory: principles and applications*, 100, 1999.
20. Prapanporn Rattana, Konstantin B Blyuss, Ken TD Eames, and Istvan Z Kiss. A class of pairwise models for epidemic dynamics on weighted networks. *Bulletin of mathematical biology*, 75(3):466–490, 2013.
21. Martin Ritchie, Luc Berthouze, and Istvan Z Kiss. Beyond clustering: Mean-field dynamics on networks with arbitrary subgraph composition. *Journal of mathematical biology*, 72(1-2):255–281, 2016.
22. Kieran J Sharkey, Carmen Fernandez, Kenton L Morgan, Edmund Peeler, Mark Thrush, James F Turnbull, and Roger G Bowers. Pair-level approximations to the spatio-temporal dynamics of epidemics on asymmetric contact networks. *Journal of mathematical biology*, 53(1):61–85, 2006.
23. Neil Sherborne, Joel C Miller, Konstantin B Blyuss, and Istvan Z Kiss. Mean-field models for non-markovian epidemics on networks. *Journal of mathematical biology*, 76(3):755–778, 2018.

24. Pieter Trapman. On analytical approaches to epidemics on networks. *Theoretical population biology*, 71(2):160–173, 2007.
25. Erik M Volz, Joel C Miller, Alison Galvani, and Lauren Ancel Meyers. Effects of heterogeneous and clustered contact patterns on infectious disease dynamics. *PLoS computational biology*, 7(6):e1002042, 2011.